MANSOURA JOURNAL OF BIOLOGY
Official Journal of Faculty of Science, Mansoura University, Egypt
E-mail: scimag@mans.edu.eg ISSN: 2974-492X

Tertiary Structure Prediction and Analysis of the Non-structural Protein 5B of Hepatitis C Virus Genotype 4a

Shaimaa Abd El-Ghany; A. A. El-Morsi; Mahmoud El-Bendary; Mustafa Neamatallah; Yehia A. Osman

Associate professor of Microbiology, Botany Department, Faculty of Science, Mansoura University

Professor of Tropical Medicine, Tropical Medicine Department, Faculty of Medicine, Mansoura University

professor of Biochemistry and Molecular Medicine, Biochemistry Department, Faculty of Medicine, Mansoura University

Professor of microbiology/Molecular Biology Botany Department, Faculty of Science, Mansoura University

Received:17/1/2022
Accepted: 1/2/2022

Abstract HCV genotype testing was an essential technique before starting treatment, it's necessary to have a better understanding of molecular epidemiology and genetic variability of HCV genotypes. NS5B is a 65-kDa RNA-dependent RNA polymerase that is substantially conserved across all HCV genotypes. It was one of seven nonstructural proteins found in HCV and is required for the virus's replication. This protein is an excellent target for the development of a pan-genotypic, direct-acting antiviral (DAA) with a high resistance barrier, so The goal of this work was to genotype HCV strains circulating in the Dakahlia Governorate by adopting direct sequencing and phylogenetic study of a portion of the HCV genome encoding NS5B in order to compare the NS5B protein in the local HCV genotype 4a with that found in other countries. Using the Jalview tool for molecular evolutionary and phylogenetic research, it was discovered that the HCV NS5B gene of the local genotype 4a detected in Egypt was closely related to their counterpart in Egypt, EF694396.1:21-309, and showed high genomic diversity of genotype 4 from other parts of the world. It also illustrated that the Egyptian isolates can be classified into four clusters. J Netpred algorithm shows that secondary structure of NS5B protein composed of seven alpha helices that were specified at 9-21 (EKDIRAEEEVYQC amino acids), 26-40 (PEARKVITALTERLY amino acids), 70-84 (NTLTCYLKATAAIRA amino acids), 108-120 (VEEDNRALRAFTE amino acids), 122-124 (MTR amino acids), 168-175 (ETPLARAA amino acids), 184-198 (VNSWLGNIIVYAPTI amino acids), and at 203-214 (VLMTHFFSILQS amino acids) and 10 β -sheets at 40-46 (YVGGPMH amino acids), 51-55 (DLCGY amino acids) , 91-94 (TMLV amino acids), 99-104 (LVVIAE amino acids), 134-140 (QPAYDLE amino acids), 142-143 (IT amino acids), 149-154 (VSVAHD amino acids), 159-164 (KVYYLT amino acids), 177(E amino acid) and at 221-233 (ALDFDMYGVTYSI amino acids). The 3-D structure prediction model for the NS5B protein of different HCV genotype 4a isolates showed 4 chains: A, B, Q and R. In conclusion, the Egyptian genotype 4a was structurally unique and should lead to development of a specific viral drug resistance mechanisms, which results in the need to devise its own antiviral medicines.

keywords: Hepatitis C virus, NS5B protein, protein structure prediction, phylogenetic analysis.

1.Introduction

There has been a 2.8% rise in the number of people who have hepatitis C virus (HCV) infections over the last decade, which has led to more than 185 million infections (3 percent of the world's population). 119 million people in the world's adult population have chronic HCV infection, and about 4 million people become infected each year. 350000 to 500000 people die each year as a result of HCV-related complications, and 350000 to 500000 people

die each year **[1,2]** About 50–80 percent of people who get HCV become chronic active cases, and they go through the stages of fibrosis, cirrhosis, hepatocellular carcinoma, and death before they can get better **[3].** China, Pakistan, Nigeria, Egypt, India, and Russia made up more than half of all infections. Nearly three-quarters of people who have been infected live in middle-income countries **[4].** In the family Flaviviridae, HCV is part of the Hepacivirus genus. One single-stranded RNA molecule of about 9.6 kb in length makes up the entire HCV genome. The coding region is made up of one large open reading frame (ORF), which is flanked by non-translated regions (NTR) at the 5' and 3' ends. The ORF is surrounded by non-translated regions (NTR). **[5].** A single polyprotein is encoded by the virus, which is divided into ten distinct proteins via cellular and viral proteases, including NS2, NS3, NS4A, NS4B, NS5A, and NS5B are all structural proteins, as are E1, E2, and p7, which are all nonstructural proteins **[6].** The RNA dependent RNA polymerase (RdRp) is formed by the nonstructural 5B (NS5B) **[7].** The 65-kD NS5B protein, which is an origin of viral replication, is found at the C-terminal region of the HCV polyprotein and shares sequence similarities with animal viruses, plant viruses and bacteriophages **[8,9,10].** According to phylogenetic and sequencing investigations of the entire viral genomes, HCV strains are divided into seven genotypes and many subtypes **[11].** HCV genotypes vary by 31–33%, while subtypes within a genotype vary by 20–25%. **[12].** HCV genotypes are dispersed globally, with genotypes one and two being the most common in West Africa, three in South Asia, four in Central Africa and the Middle East, five in Southern Africa, and six in South East Asia **[13,14]** and genotype seven in Congo **[15].** Around 20% of all instances of chronic HCV infection in the globe are caused by the genotype four strain **[16].** Amplification of HCV subgenomic regions (core, E1, and NS5) followed by sequence analysis is now widely accepted as the gold standard for HCV genotyping and is commonly used **[17,18].** In comparison to all other HCV genes, the NS5B gene possesses the largest phylogenetic signal **[19].** So NS5B is used to differentiate between various isolates of the same subtypes as it is more stable over time. Due to its stability reduce the minimum risk of two different strains when comparing between them rather than identifying the existence of these strains of the same origin that vary from the typical HCV genome when the immune system is activated **[20].**

In Egypt, where an estimated 15% of the population may have chronic hepatitis C, than 90% of them are HCV genotype four **[11,21].** Furthermore, In Egypt, subtype 4a has been identified as the most common, while infection with genotype 1 never exceeded 10% **[22,23,24].** HCV genotyping is significant not only for predicting the efficacy of antiviral therapy, but also for establishing how long it should be taken **[25].** Moreover, it's a good way to learn more about the virus's epidemiological and virological properties **[26].** Therefore the present study was aimed to investigate and throw more lights on the NS5B of HCV genome which plays an essential role in initiation of viral RNA replication where it was confined to HCV patients from Dakahlia governorate, Egypt through a project funded by the Science and Technology Development Fund (STDF) It aimed to genotype HCV strains spreading in Dakahlia Governorate and predict the three dimensional (3-D) structures of virus NS5B protein to elucidate their functions and integrative understanding of viral processes.

2. Materials and methods

1-Sampling

Hundred samples with their clinical and biochemical profiles were collected from clinics of the Tropical Medicine Department at Mansoura University, from 2014 to 2016. Serostatus of each participant was determined by testing Utilizing a commercial enzyme-linked immunosorbent assay (ELISA) kit (DIA, PRO), test for anti-HCV antibodies (anti HCV Ab).

2-Viral RNA extraction and Sequencing

In accordance with the manufacturer's recommendations, viral RNA was isolated from serum samples using a Qiagen RNA extraction kit (Qiamp® RNA Mini kit, Qiagen, German). According to the manufacturer's instructions, the isolated The Quant One Step RT-PCR Kit was used to convert viral RNA to cDNA and amplify it (Tiangen). Non-structural HCV

genome sequencing was used to identify HCV genotypes (NS5B) viral genes directly with the primers sense 5' TTCTCRTATGAYACCCGCTGYTTTGA-3' and antisense 5'-TACCTVGTCATAGCCTCCGTGAA-3 ' **[27].**

3- Determining the conserved regions of NS5B

Multiple sequence alignments of the NS5B gene and protein of the genotype 4a, was conducted using Jalview 2.10.2 (www.jalview.org/) program **[28].** in which sequences can be aligned by means of a variety of algorithms provided by JABA web services, together with Clustal W, Muscle, MAFFT, ProbCons, T-COFFEE and Clustal Omega **[29].** While the most widely used Clustal Omega was used for multiple sequence alignment of the DNA showing consensus sequence from the HCV database.

4- Molecular evolutionary and phylogenetic analysis

The Jalview program 2.10.2 was used for construction of phylogenetic tree of all HCV NS5B sequences (nucleotide and amino acids) of the genotype [4a] from different countries to demonstrate the evolutionary relationship among them. The phylogenetic tree had been established using Neighbor-Joining (NJ) method on the Jalview 2.10.2 program **[28]** Saitou and his colleagues developed neighbour joining, a bottom-up clustering approach for rebuilding phylogenetic trees using evolutionary distance data by **[30].** The approach, which is often used for trees based on DNA or protein sequence data, needs knowledge about the distance between each pair of taxa (e.g., species or sequences) in order to create the tree **[31].** The phylogenetic tree was built using HCV sequences from other HCV isolates collected from the HCV database and GenBank and also making Sequence locator. Removing vector sequences of obtained NS5B sequences using VecScreen. Obtaining restriction maps using NEB cutter (Restriction Mapping) and Estimation of % G~C Count of all NS5B sequences.

5- Prediction of NS5B Secondary structure using PSIPRED software

The consensus sequence of NS5B protein obtained from the previous alignments was used in FASTA format as input for the PSI-blast based secondary structure prediction (PSIPRED) server (http://bioinf.cs.ucl.ac.uk/psipred) which is a simple and precise secondary structure prediction method, combining two feed-forward neural networks which carry out the analysis on the output of PSI-BLAST (Position Specific Iterated – BLAST). For the most accurate prediction, All 25 downloadable NS5B protein sequences of genotype 4a from various countries, including Egypt, were used to create a consensus sequence. The secondary structure prediction was obtained using Jpred4 server **[32].**(http://www.compbio.dundee.ac.uk/jpred4) JPred4 is the most current version of the popular JPred protein secondary structure prediction service, which delivers predictions using the JNet algorithm, one of the most exact techniques for secondary structure prediction, as well as solvent accessibility and coiled-coil regions. In addition, the protein prediction website was utilised to identify the exposed and buried portions of the NS5B protein. Each amino acid is classified by this server into one of four groups (all-alpha, all-beta, alpha-beta and mixed all others).

6- Prediction of NS5B protein tertiary structure

This prediction was achieved using, as inputs it was used profile-profile alignments and PSIPRED to predict secondary structure by selecting the most accurate one with the highest confidence which obtained from the previous step and colored by green. Then the result was converted to another database in the protein data bank (PDB) called the PDBsum which is a graphic database that that gives an insight of the interior of each 3D structure in the PDB It shows the molecules that make up the structure (protein chains, DNA, ligands, and metal ions) as well as schematic depictions of their relationships. It also can extensively make use of the freely obtainable RasMol molecular graphics program for viewing the molecules and their interactions in 3D such as 3Djmol. The Jmol server was used to predict the 3D structure of the NS5B protein **[33].** which is an open-source Java viewer for chemical structures in 3D (http://www.jmol.org/). All 25

downloaded genotype 4a protein sequences from multiple countries, including Egypt, were used to create a consensus sequence. The NS5B location of the HCV genome is 389 nucleotides long and covers the central segment of the NS5B gene. In ten of the highest viral load positive samples, The HCV NS5B gene was successfully sequenced directly and phylogenetically analysed. To obtain different genotypes specific for NS5B of HCV, partial NS5B coding sequences derived from our samples in FASTA format were aligned with GenBank sequences (http://www.blast.ncbi.nlm.nih.gov) and to demonstrate the effect of geographical distribution on genotype genetic diversity, we aligned them with GenBank sequences (http://www.blast.ncbi.nlm.nih.gov).

3. Results and Discussion

1-Multiple sequence alignment and phylogenetic analysis of HCV-NS5B (Sequence Analysis)

The nucleotide sequence of HCV–NS5B was compared on the levels of DNA Figure (1) and deduced amino acids Figure (2) for contig DAk1 with different sequence isolated

internationally from different location in the world. Multiple sequence alignment used to confirm authenticity of the target NS5B gene and to compare between the isolated sequences and other isolated sequences from different regions.

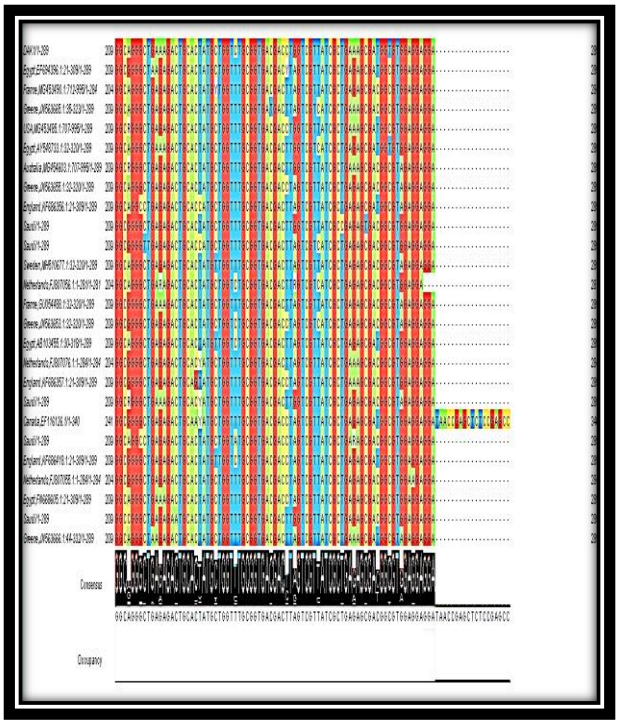

Fig1. Multiple sequence alignment of NS5B DNA Contig sequences (nucleotides) with NS5B from Egypt and other different geographical Viruses from other countries are seen in Jalview, with a conserved segment and consensus sequence identified.

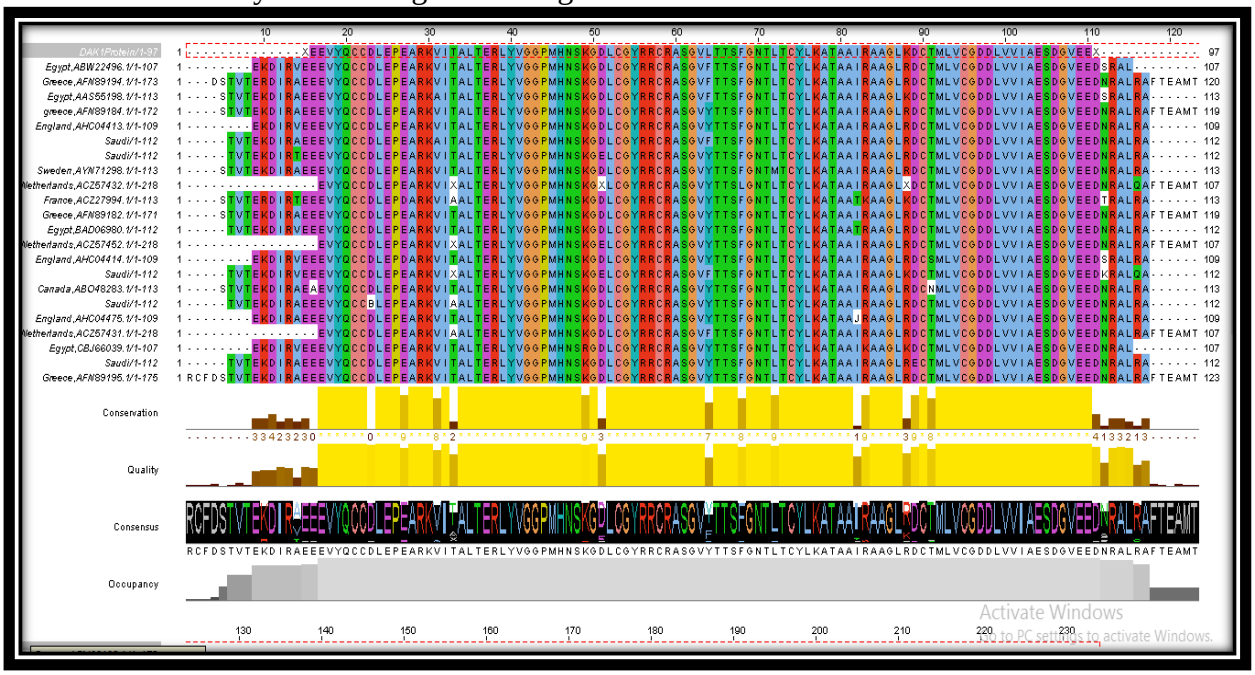

Fig 2. Multiple sequence alignment of NS5B protein sequences (amino acids) of genotype 4a using Jalview and showing highly similar conserved domains and consensus sequence.

2- Molecular evolutionary and phylogenetic analysis

The Jalview program was used for construction of phylogenetic tree of all HCV

NS5B sequences of the genotype [4a] from different countries to demonstrate the evolutionary relationship among them. The phylogenetic tree was built using the Neighbor-Joining (NJ) technique **[30].** Phylogenetic analysis of NS5B was done for both DNA as shown in figure (3) and for protein as shown in figure (4).

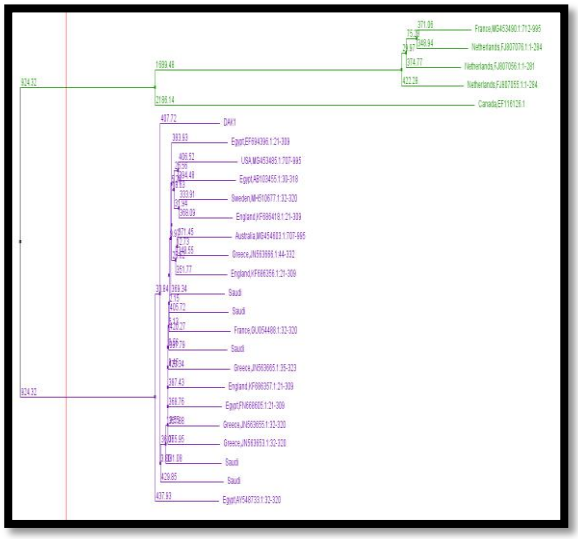

Fig3. Phylogenetic analysis based on the nucleotide sequences showing genetic relationship between NS5B DNA sequences (4a) of the selected isolates from Gen Bank and the amplified segment (contig).

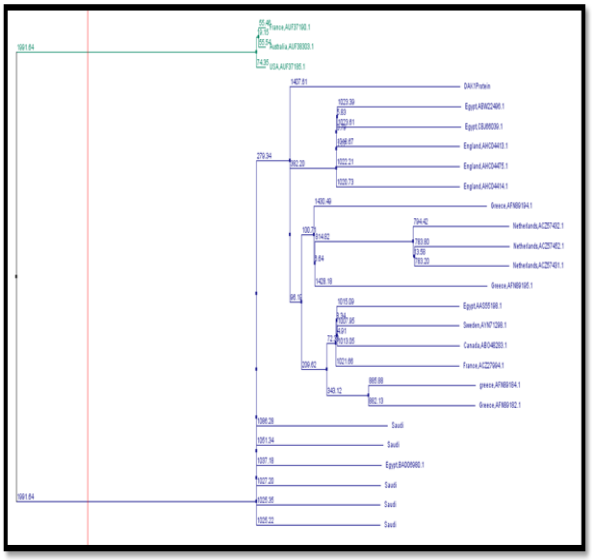

Fig 4. Phylogenetic tree of NS5B protein sequences(4a) from different countries including Egypt, downloaded from Gen Bank showing degree of kinship and the effect of geoghraphical distribution on it.

3- Secondary structure

Prediction Secondary structure prediction of NS5B using PSIPRED

An accurate secondary structure of NS5B protein genotype 4 subtype (a) was predicted using PSIPRED and showing 7 alpha helices and also 10 β -sheets (strands) with the rest of amino acids showing coils as illustrated in figure (5).

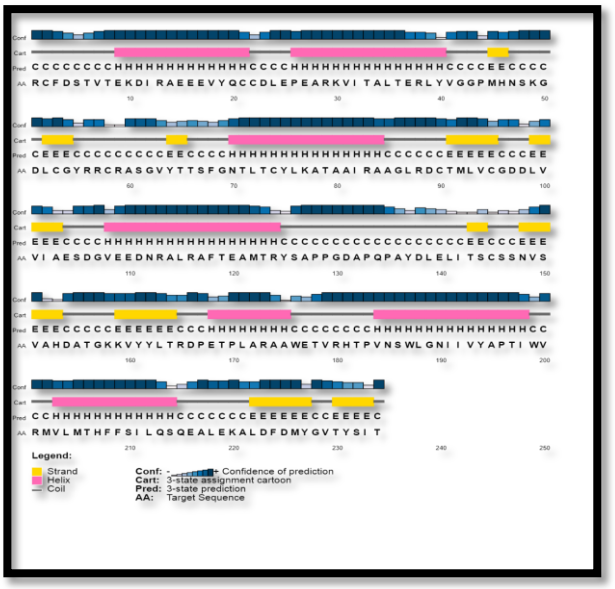

Fig 5. The predicted protein secondary structure of NS5B protein showing helix, strand and coils produced by PSIPREDView-a Java visualization tool.

To improve the precision with which secondary structure can be predicted, The J.Pred 4 programme for secondary structure prediction used the consensus sequence of all downloaded NS5B protein sequences of genotype 4a from various countries as input. Figure 6 depicts the secondary structure prediction of the consensus NS5B protein of genotype 4a, which contains 234 amino acids.

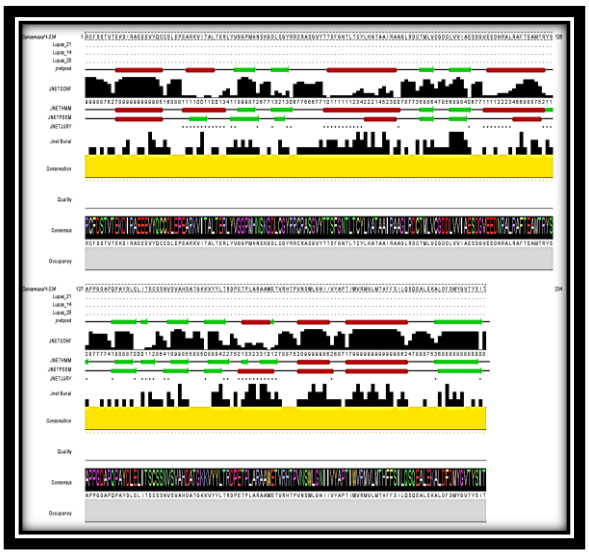

Fig 6. The Secondary structure prediction for NS5B protein (consisting of 234 amino acids)

of HCV genotype 4, subtype a isolates from different countries in the world (using JPred4). The results show that it has 7 alpha helix and 10 beta sheets.

4- 3D Structure Prediction of NS5B protein

Initial models are built, developed, and reviewed before the highest quality one is selected to produce the 3-D model. Jmol server was used to estimate the 3D structure of the consensus sequence of all downloaded NS5B protein genotype 4a from various nations, and the resulting structure is illustrated in Figure (7).

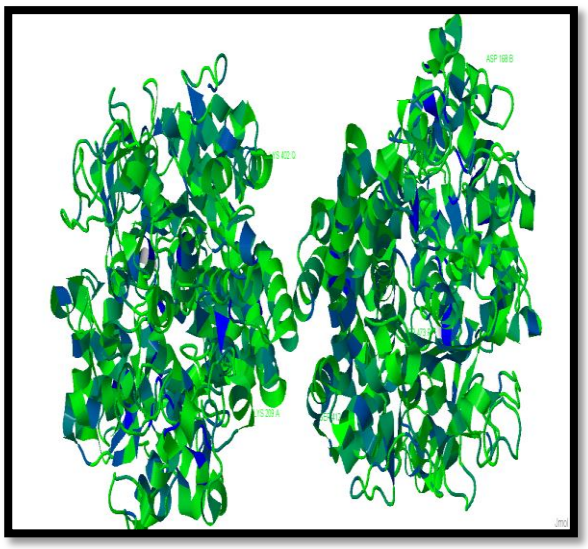

Fig 7. 3D (tertiary) structure prediction model for the NS5B protein of different HCV isolates (using J. mol) showing 4 chains (A, B, Q and R).

The entire economic burden of HCV in Egypt was projected to be $3.81 billion USD, accounting for around 1.4 percent of the country's total GDP (GDP) as estimated by **[34].** The diseases reach 700 million dollars per year, accounting for nearly 4% of Egypt's total health-care spending, reflecting the significant health and financial costs of HCV and its consequences **[35].** The NS5B gene's nucleotide sequencing is the gold standard for precise HCV genotyping **[36].** HCV genotyping utilising NS5B nucleotide Sequence analysis is quite effective, allows for effective subtype identification, and is a useful method for studying HCV molecular epidemiology **[27].** All samples in the current investigation were genotype 4 with subtypes 4a based on HCV NS5B gene direct sequencing and phylogenetic analysis fully agreement with previous studies of **[37].** that revealed, The majority of

Egyptian patients had genotype 4 infection, with subtype 4a seems to be the most common and partially similarities with other investigations, from the Alexandria area, revealed that 78 percent of the isolates were HCV genotype 4a variations, with the remaining detected variants being 4m (11 percent), 4o (5.5 percent), 4n (2.7 percent), and 4p (2.7 percent) (23) **[38]**, as well as that of Ismailia where, 15 patients (78.9%) were subtype 4a, two (10.5%) were subtype 1g, and one (5.2%) was linked with subtype 4o **[22].** Bioinformatics methods are especially important in the HCV field because a high percentage of HCV research involves sequence analysis of clinical specimens. This research reflects that Technical restrictions, like as cell culture techniques that only permit replication of a single variety, have hampered HCV research. and there is no tractable small animal model; detection of HCV RNA by in situ is difficult due to low levels of viral replication. Thus, *in silico* methods are extremely important in HCV virology. We have used bioinformatics tools to facilitate sequence and statistical analysis and applied these tools to studies of the HCV genome **[39].** On the present study HCV antibodies were detected by ELISA kit (DIA.PRO) in 54 serum patients from 100 samples with values ranged from 0.11 to 5.3 similar work was obtained by **[40].**Moreover, in order to select the highest viral load samples (*i.e.* viraemic infection) For the 54 HCV strains, a quantitative quantification of HCV-RNA in serum was found using an RT-PCR-based test. seropositive samples the detected results were between (8980 –3,672,120 IU/ml) with low, moderate and high titers. Similar work was done by **[41].**The HCV-RNA was extracted from the only 10 samples of the highest viral load positive samples of HCV and used as a template of RT-PCR to amplify cDNA that used as a target for isolation of the NS5B fragment to give PCR product with 389 bp using specific primer similar work was suggested by **[27].** The results reported here in the current study direct provision of healthcare expenses of HCV-related concerning the oligonucleotide sequence of NS5B fragments from DAK1 Contig showing query length 289

with location on reference sequence H77 from nucleotide number 8308 to 8596. The similarity between the sequence under investigation and 25 overseas isolates of HCV-NS5B was determined. Results showed that, the present isolate appeared (92%) identity on the levels of DNA sequences as well as amino acids with the isolate of Egypt. Molecular evolutionary and phylogenetic analysis using Jalview program showed that the HCV NS5B gene of the local genotype 4a detected in Egypt was closely related to their counterpart in Egypt, EF694396.1:21-309. And showed high genomic diversity of genotype 4 from other parts of the world. It also illustrated that the Egyptian isolates can be classified into four clusters.

To gain insights into the biological importance of proteins, DISOPRED3 can also predict binding sites within disordered regions of protein, which helps in discovering evolutionary conserved patterns, positional information and amino acid sequence composition of putative disordered regions **[42].**

The experimental results reported here predicted the secondary structure of NS5B protein using PSIPRED and showing 8 alpha helices with the rest of amino acids showing coils. Furthermore, the 3-D structure prediction model for the NS5B protein of different HCV genotype 4a isolates showed 4 chains: A, B, Q and R. This work too close to that reported by **[33].** In conclusion, The HCV NS5B RdRp element is an invariant element in RNA virus polymerases that is highly conserved across all HCV genotypes. As a result, the HCV RdRp provides an attractive target for developing a pan-genotypic, direct-acting-antiviral (DAA) with a high resistance barrier as the Egyptian genotype 4a was structurally unique and should lead to development of a specific viral drug resistance mechanisms, which results in the need to devise its own antiviral medicines.

4. References

1 Petruzziello, A.; Marigliano, S.; Loquercio, G.; Cozzolino, A. and Cacciapuoti, C. (2016). Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. World *journal of gastroenterology*,**22(34):** 7824.

2 WHO. (2016.) Hepatitis C fact ., sheet no. 164.

3 Stanaway, J.; Flaxman, A.; Naghavi, M.; Fitzmaurice, C.; Vos, T.; Abubakar, I.; Abu-Raddad,L.; Assadi, R.; Bhala, N. and Cowie, B. (2016). The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. The Lancet,388(10049): 1081-1088.

4 Hanafiah, K.; Groeger, J.; Flaxman, A. and Wiersma , S. (2013). Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. Hepatology,**57(4)**: 1333-1342.

5 Moradpour, D.; Penin, F. and Rice , C. (2007). Replication of hepatitis C virus. Nature reviews microbiology, **5(6)**: 453-463.

6 Quinkert, D.; Bartenschlager, R. and Lohmann, V. (2005). Quantitative analysis of the hepatitis C virus replication complex. *Journal of virology*,**79(21**): 13594-13605.

7 Waheed, Y.; Saeed ,U.; Anjum ,S.; Afzal , M. S. and Ashraf ,M. (2012). Development of global consensus sequence and analysis of highly conserved domains of the HCV NS5B prote in. Hepatitis monthly,**12(9)**.

8 Aligeti, M.; Roder, A. and Horner, S. (2015). Cooperation between the hepatitis C virus p7 and NS5B proteins enhances virion infectivity. *Journal of virology*, **89(22)**: 11523-11533.

9 Poch, O.; Sauvaget, I.; Delarue, M. and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *The EMBO journal*, **8(12)**: 3867-3874.

10 Müller, R.; Poch, O.; Delarue, M.; Bishop, D. and Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *Journal of General Virology*, **75(6)**: 1345-1352

11 Smith, D.; Bukh, J.; Kuiken, C.; Muerhoff, A.; Rice, C.; Stapleton, J. and Simmonds, P. (2014). Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated

criteria and genotype assignment web resource. Hepatology,**59(1)**: 318-327.

12 Simmonds, P.; Bukh, J.; Combet, C.; Deléage, G.; Enomoto, N.; Halfon, S.; Inchauspé , G.; Kuiken , C. and Maertens , G. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology, **42(4)**: 962-973.

13 Foote, B. S.; Spooner, L. M. and Belliveau, P. P. (2011). Boceprevir: a protease inhibitor for the treatment of chronic hepatitis C. Annals of Pharmacotherapy, **45(9)**: 1085-1093.

14 Labarrere, C.; Woods, J.; Hardin, J.; Campana, G.; Ortiz, M.; Jaeger, B.; Reichart, B.; Bonnin, J.; Currin, A. and Cosgrove, S. (2011). Early prediction of cardiac allograft vasculopathy and heart transplant failure. *American Journal of Transplantation*, **11(3)**: 528-535.

15 Murphy, D.; Sablon, E.; Chamberland, J.; Fournier, E.; Dandavino, R. and Tremblay , C. (2015). Hepatitis C virus genotype 7, a new genotype originating from central *Africa. Journal of clinical microbiology*, **53(3)**: 967-972.

16 Khattab, M.; Ferenci, P.; Hadziyannis, S.; Colombo , M.; Manns, M.; Almasio, P.; Esteban, R.; Abdo , A.; Harrison, S. and Ibrahim, N. (2011). Management of hepatitis C virus genotype 4: recommendations of an international expert panel. *Journal of hepatology*, **54(6)**: 1250-1262.

17 Simmonds, P.; Smith, D. McOmish, F. Yap, P. Kolberg, J. Urdea, M. and Holmes ,E. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *Journal of General Virology*, **75(5)**: 1053-1061.

18 Nakatani, S. M.; Santos, C. A. Riediger, I. N. Krieger, M. A. Duarte, C. A. do Carmo Debur, M. Carrilho, F. J. and Ono, S. K. (2011). Comparative performance evaluation of hepatitis C virus genotyping based on the 5'untranslated region versus partial sequencing of the NS5B region of brazilian patients with chronic hepatitis C. *Virology journal*, **8(1)**: 1-5.

19 Pasquier, C.; Njouom, R. Ayouba, A. Dubois, M. Sartre, M. T. Vessière, A. Timba, I. Thonnon, J. Izopet, J. and Nerrienet, E. (2005). Distribution and heterogeneity of hepatitis C genotypes in hepatitis patients in Cameroon. *Journal of medical virology*, **77(3):** 390-398.

20 Baclig, M. O.; Chan, V. F. Ramos, J. D. A. Gopez-Cervantes, J. and Natividad, F. F. (2010). Correlation of the 5′ untranslated region (5′ UTR) and non-structural 5B (NS5B) nucleotide sequences in hepatitis C virus subtyping. *International journal of molecular epidemiology and genetics*, **1(3):** 236.

21 Guerra, J.; Garenne, M.; Mohamed, M. and Fontanet, A. (2012). HCV burden of infection in Egypt: results from a nationwide survey. *Journal of viral hepatitis*, **19(8)**: 560-567.

22 Youssef, A.; Yano, Y.; Utsumi, T.; Abd El-alah, E.; Abd El-Hameed, A.; Serwah, A. and Hayashi, Y. (2009). Molecular epidemiological study of hepatitis viruses in Ismailia, Egypt. Intervirology,**52(3):** 123-131.

23 Fakhr, A.; Pourkarim , M.; Maes, P.; Atta, A. H.; Marei, A.; Azab,M. and Van Ranst, M. (2013). Hepatitis C virus NS5B sequence-based genotyping analysis of patients from the Sharkia Governorate, Egypt. Hepatitis Monthly, **13(12).**

24 Ray, S.; Arthur, R.; Carella, A.; Bukh, J. and Thomas, D. (2000). Genetic epidemiology of hepatitis C virus throughout Egypt. *The Journal of infectious diseases*, **182(3)**: 698-707

25 Cavalcante, L. N. and Lyra, A. C. (2015). Predictive factors associated with hepatitis C antiviral therapy response. World *journal of hepatology*, **7(12)**: 1617.

26 Rajhi, M.; Ghedira, K.; Chouikha, A.; Djebbi, A.; Cheikh, I.; Ben Yahia, A.; Sadraoui, A.; Hammami, W.; Azouz, M. and Ben Mami, N. (2016). Phylogenetic analysis and epidemic history of hepatitis C Virus genotype 2 in Tunisia, North Africa. PLOS one, **11(4):** e0153761.

27 Murphy, D. G.; Willems, B. Deschênes, M. Hilzenrat, N. Mousseau, R. and Sabbah, S. (2007). Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5′ untranslated

region sequences. *Journal of clinical microbiology*, **45(4)**: 1102-1112.

28 Waterhouse, A. M.; Procter, J. B. Martin, D. M. Clamp, M. and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, **25(9):** 1189-1191.

29 Sievers, F.; Wilm, A. Dineen, D. Gibson, T. J. Karplus, K. Li, W. Lopez, R. McWilliam, H. Remmert, M. and Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology**, 7(1)**: 539.

30 Saitou, N. and Neim M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, **4(4)**: 406-425.

31 Xavier, D. (2010). Sequence-Based Analysis of Bacterial Population Structures. In:Bacterial Population Genetics in Infectious Disease .Robinson, D.; Falush, D. and Edward, J. F. (eds.). John Wiley and Sons, 46–47.

32 Alexey, D.; Christian, C.; James, P. and Geoffrey, J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res., 43: 389–394.Abdurakhmonov, I. Y. (2016). Bioinformatics: basics, development, and future, InTech.

33 Herraez, A. (2006). Biomolecules in the computer: Jmol to the rescue. Biochemistry and Molecular Biology Education, **34(4)**: 255-261.

34 Central Agency for Public Mobilization and Statistics (CAPMAS) (2017). Statistical Yearbook – National Accounts: Expenditure on GDP at Market 12/13 – 14/15. Availablefrom: http://www.capmas.gov.eg/Pages/StaticPages.aspx?page_ido5034 .Accessed April 13, 2017

35 World Bank(2017). Health Expenditure, Total (% of GDP). Available from: http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS. Accessed April 13, 2017.

36 Chaiwong, S. and Sistayanarain, A. (2016). Characterization of partial NS5B region among hepatitis C virus genotype 6 subtypes isolated from Thai blood donors. *Journal of medical virology*., **88(10)**: 1785-1790.

37 Kamal, S. M. and Nasser, I. A. (2008). Hepatitis C genotype 4: What we know and what we don't yet know. Hepatology., **47(4)**: 1371-1383.

38 Genovese, D.; Dettori, S., Argentini, C.; Villano, U.; Chionne, P.; Angelico, M. and Rapicetta, M. (2005). Molecular epidemiology of hepatitis C virus genotype 4 isolates in Egypt and analysis of the variability of envelope proteins E1 and E2 in patients with chronic hepatitis. *Journal of clinical microbiology*, **43(4)**: 1902-1909.

39 Zheng, N., Wang, K., Zhan, W. and Deng, L. (2019). Targeting virus-host protein interactions: Feature extraction and machine learning approaches. Current drug metabolism., **20(3)**: 177-184.

40 Alzahrani, A. J. and Obeid, O. E. (2004). Detection of hepatitis C virus core antigen in blood donors using a new enzyme immunoassay. *Journal of family & community medicine*, 11(3): 103.

41 Cook, L.; Wing, N.; Arthur, B.; Lawrence, C. and Keith, R. (2004). Use of the MagNA pure LC automated nucleic acid extraction system followed by real-time reverse transcription-PCR for ultrasensitive quantitation of hepatitis C virus RNA. *Journal of Clinical Microbiology*, **42 (9):** 4130-4136.

42 Jones, D. T. and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics., **31(6):** 857-863.